Micro-work, artificial intelligence and the automotive industry

Paola Tubaro, CNRS*

Antonio A. Casilli, Telecom ParisTech

Abstract

This paper delves into the human factors in the "back-office" of artificial intelligence and of its data-intensive algorithmic underpinnings. We show that the production of AI is a labor-intensive process, which particularly needs the little-qualified, inconspicuous and low-paid contribution of "micro-workers" who annotate, tag, label, correct and sort the data that help to train and test smart solutions. We illustrate these ideas in the high-profile case of the automotive industry, one of the largest clients of digital data-related micro-working services, notably for the development of autonomous and connected cars. This case demonstrates how micro-work has a place in long supply chains, where tech companies compete with more traditional industry players. Our analysis indicates that the need for micro-work is not a transitory, but a structural one, bound to accompany the further development of the sector; and that its provision involves workers in different geographical and linguistic areas, requiring the joint study of multiple platforms operating at both global and local levels.

Keywords

Artificial intelligence, micro-work, automotive industry, digital platform economy, organization of work.

Notes

This is a pre-print of an article published in *Journal of Industrial and Business Economics*. The final authenticated version is available online at: https://doi.org/10.1007/s40812-019-00121-1

* Corresponding author. Laboratoire de Recherche en Informatique, Bâtiment 660 Université Paris Sud, 91405 Orsay Cedex, France, paola.tubaro@lri.fr

1. Introduction

How does the current, spectacular development of artificial intelligence (AI) affect labor and work organization? Recent attempts to address this complex question revamp the centuries-old conjecture that machines may ultimately supersede human contribution (Ricardo 1951 [1821]). News media, think tanks, and policymakers have echoed the dire job loss predictions of a few prospective studies (Brynjolfsson and McAfee 2014; Frey and Osborne 2017), although historical evidence suggests that dynamics of complementarity rather than substitution between labor and machinery might also be in place (Autor 2015).

Moving away from the highly speculative nature of these debates, we endeavor to gain new insight by looking at the concrete production processes of AI, as they are implemented today. We claim that these processes are – and will increasingly be – labor-intensive. If our stance refutes popular pessimistic scenarios of mass and persistent technological unemployment, we do not paint a rosy picture either. Indeed AI production not only relies on the human capital of highly skilled, well-paid engineers and scientists; but more subtly, it is predicated on the less visible contribution of little qualified, low-paid people who toil to ensure the "last mile" (Gray and Suri 2017). By lifting the veil on them, we highlight how their activities and roles take forward, in today's digitized context, a deeply-rooted tendency to reduce many laborers to a reserve of undifferentiated parts, mere add-ons to machines that require limited skills (Braverman 1974).

This inconspicuous, low-profile human activity behind AI is known as "micro-work", a notion that encompasses a range of small, fragmented tasks performed remotely online by large numbers of providers. Such tasks consist, for example, in identifying objects on a photograph, sorting items in a list, answering simple questions, transcribing or copying short texts, or adding labels to images. Micro-work is sometimes referred to as "crowd-work" because the mundane, repetitive, and atomized nature of this platform-mediated activity makes operators substitutable with each other: a job is not given to an identified sub-contractor but to an anonymous, undifferentiated multitude. Although each single task may be quick-to-do and unchallenging, the outputs of myriad executors add up to constitute large and valuable digital data resources which contribute to AI production: as will be discussed at length later, they help training computer vision models, operating voice-activated connected objects, providing targeted recommendations. Micro-work is a by-product of the "datafication" of our societies (Bastin and Tubaro 2018), whereby more data and increasing computational capacity have brought AI research to the fore and in turn, further raised the demand for ever-larger, ever-better data assets. It is also an instance of the "commodification" of data in our society and the ensuing inequalities, whereby a handful of powerful corporations own, control and extract value from data (Fuchs 2019).

Micro-workers are recruited through dedicated platforms, the most famous of which is Amazon Mechanical Turk, with a growing number of competitors worldwide such as Microworkers, Clickworker, Appen and Lionbridge. With various arrangements, they act as intermediaries in two-sided markets (Rochet and Tirole 2003) where clients are AI (and other data-hungry) companies on the one side, and workers on the other. With rare exceptions, workers are not formally employees, but independent contractors paid by piece-rate, with varying levels of activity and engagement. Part of the sector consists in services owned by tech giants who are their exclusive clients. This is the case of UHRS (Universal Human Relevance System) controlled by Microsoft, and RaterHub by Google. Micro-work involves laborers in over seventy countries (Berg et al 2018; Kuek et al. 2015), with many contributors in emerging economies such as India, Indonesia and the Philippines, and a recent surge in participation from crisis-affected areas such as Venezuela (Schmidt 2019).

Amazon's candid definition of micro-work as "artificial artificial intelligence" stresses its key functions: though often little qualified, these humans help AI get past those tasks and activities that it cannot solve effectively and/or efficiently. Whether or not the gap between what AI systems claim to achieve and what they need humans to achieve is bridgeable, is an open debate opposing those portraying micro-work as a transitional phenomenon and those defining it as a structural, intrinsic driver of today's AI growth. Its role seems to be changing over time: born to supplement AI when it fails (Irani 2015), micro-work has progressively moved to offering support for AI development. Ekbia and Nardi (2017) see it as a form of "heteromation", a mode of division of labor between humans and machines that reverses the myth of an ideal automation capable of freeing people from the need to work. Its very existence is evidence that, whether or not AI might displace employment in future, it is already producing an effect on work here and now, shifting groups of laborers into unrecognized, "silent" (Star and Strauss 1999) albeit essential activities at its margin. We thus see micro-work as the prism through which we can tackle the place of humans in the design and deployment of AI in today's data economy.

2. The automotive industry as a case study

We illustrate these ideas in the high-profile case of the automotive industry, which has become one of the largest clients of digital data-related micro-working services, notably for the development of autonomous vehicles and of connected cars. This case best exemplifies how micro-tasks integrate long supply chains, no longer limited to digital industries (Ohnemus 2016), but reaching out to core economic sectors, whereby major tech companies such as Google and Uber now compete with established car manufacturers worldwide. With its very sizeable demand and unprecedentedly high quality requirements (to ensure safe traffic circulation), the automotive industry has lately transformed the whole micro-work landscape. There is a form of dual dependency in that automobile producers need micro-work service providers to sustain the technological developments that keep them competitive against their peers; but they are also quasi-oligopsonists who can dictate exact specifications to suppliers, play them against one another, and pass off to them many risks, notably of variations in demand. The result is accelerated globalization, proliferation of specialized (rather than generalist) platforms (Schmidt 2019), and development of specific AI offers by all players.

Of note, AI-based innovations in the automotive industry are not all meant to be labor-saving. Drivers do not disappear in connected cars where (quite on the contrary) the goal is to provide an improved experience, notably through hands-off speech interfaces and a range of assistance services. Even autonomous vehicles still require human "safety drivers" who routinely provide feedback to engineering teams (and are expected to take control whenever necessary). These human inputs are unlikely to disappear entirely, even though contractual arrangements might change (for example, companies may eventually replace paid safety drivers by assigning extra responsibilities to passengers). Likewise, we will show in what follows that the need for off-the-road, data-related micro-work is a structural feature of today's AI production processes – thus, it is here to stay.

3. Methodology

To develop our argument, we leverage data from a detailed inventory of micro-working platforms that we have built between June 2017 and March 2019. The focus of our study is France, a country that is investing heavily in the development of AI; yet the market for micro-working services crosses national borders, especially as most tasks can be performed remotely online and rarely require physical presence in a given place. Therefore, our inventory also includes platforms that are based in France but outsource work to providers overseas (especially in French-speaking African countries such as the Ivory Coast, Madagascar and Tunisia), and international platforms that have operations (with clients and/or workers) in France. We select from it a sample of 11 platforms that have AI as a specific selling point, even though some of them also offer services for a wider range of corporate needs such as marketing and sales. Table 1 summarizes their characteristics.

Platform	Offer	Headquarters	Foundation	Notes
Amazon Mechanical Turk	Generalist	Seattle (USA)	2005	Pioneer in the field.
Pactera	Generalist	Dalian (China)	1995	General IT service company; gives access to other platforms.
Clixsense	Generalist	Hampstead (USA)	2007	Gives access to the tasks of Figure Eight and other data vendors.
Clickworker	Generalist (with specific AI offer)	Essen (Germany)	2005	Gives access to UHRS (Microsoft).
Microworkers	Generalist (with specific AI offer)	Dallas (USA)	2009	
Lionbridge	Generalist (with specific AI offer)	Waltham (USA)	1996	Originally a translation service; has subsequently diversified into data services. Owns Gengo.ai, a specialist company.
Appen	Specialized AI	Charswood (Australia)	1996	Originally a more generalist IT service, recently re-focused on AI.
Figure Eight	Specialized AI	San Francisco (USA)	2007	Formerly called Crowdflower; part of Appen since 2019. Recruits workers through Clixsense.
Mighty AI	Specialized AI	Seattle (USA)	2014	A service for companies; the related micro-work platform is Spare5.
Wirk	Specialized AI	Paris (France)	2018	A service for companies; the related (generalist) micro-work platform is Foule Factory, founded in 2014 and recruiting workers only in France.
IsAHit	Generalist (with specific AI offer)	Paris (France)	2016	Access reserved to workers in (mainly French-speaking) African countries.

Table 1 : List of micro-working services and platforms included in the study. All have an offer of AI-related services, more or less explicitly, and have operations (clients and/or workers) in France.

For each of these platforms, we use information retrieved from their websites and communication packages, press articles concerning their businesses, and further data from aggregator services such as CrunchBase. We also use insights from in-depth interviews that we conducted in 2017-2018 with 92 micro-workers, platform operators, and clients of these services, most of them based in France.

Our selection comprises multiple platforms because the most intensely researched one, Mechanical Turk, is not representative of current trends: it is one of the few that does not explicitly advertise AI services to its clients, although it is clear from the literature that it does offer them (Irani 2016). Amazon's creation is not representative of workers' diversity either, in that a conspicuous part of them are in fact US residents (Difallah et al. 2018); specifically, very few are from France. Mechanical Turk operates primarily as an intermediary between clients and workers, without offering much in terms of workforce management and quality control – which need to be ensured by clients.

Of the other platforms listed here, Pactera and Clixsense are multi-activity IT companies that offer other services together with micro-work. They act as recruiters or dispatchers for other micro-working platforms, which is why their offer for the automobile sector is less apparent in their communication packages. For example, Clixsense provides workers with access to the tasks of Figure Eight (formerly called Crowdflower and, since 2019, part of the Australian giant Appen), which deals only with client companies and thus has no direct contact with workers. All other platforms have an explicit AI offer, based on use of micro-work. Some of them like Figure Eight, Mighty AI, Clickworker, Appen, IsAHit, and to a lesser extent Wirk, explicitly advertise their AI services to automotive companies; Lionbridge does so through Gengo.ai, a Japanese platform which it acquired in 2019.

The business models of these platforms differ. While some (like Mechanical Turk) function as textbook cases of intermediaries in two-sided markets, others (like Appen and Pactera) have multi-layer structures, whereby different platforms (or companies) take different roles. One recruits workers; another manages registration, sign-on, allocation of tasks and day-to-day supervision; a third one provides the user interface to execute tasks; yet another liaises with clients. Bigger or more demanding client companies value these more complex structures which provide them with extra services (from technical tools to workforce management) and add a degree of opaqueness that shields them from the gaze of competitors. Although the details of the underlying commercial relationships are undisclosed, these platforms often attract large orders and can thus offer less volatile earning opportunities to workers.

Wirk and IsAHit are French companies, of relatively small size compared to the others, which are all international and rather large. Created in 2018, Wirk provides data services to companies, with a prominent AI offer. It is part of the same company that manages Foule Factory (from “foule”, the French word for “crowd”), a micro-work platform founded in 2014, which recruits contributors only from among residents of France, and is a supplier to Wirk. Foule Factory is a generalist service, where micro-workers perform services for varied needs within the digital economy, in addition to AI development. IsAHit is a Paris-based company practicing “impact sourcing” along the lines of the American Samasource: its clients are mostly French while workers are recruited in low-income African countries where remunerations from micro-work compare favorably with local averages. It targets women (considered as particularly interested in opportunities to earn income from home) in French-speaking African countries, although it counts a small number of men among its workers and has recently expanded into English-speaking Africa. With regard to clients, IsAHit is not a mere intermediary but a contractor, which manages all steps of the execution of work on their behalf.

The interest of including them in our selection goes beyond the intrinsic relevance of the chosen country case study. We expect international platforms to attract only specific segments of the demand and supply of micro-work, notably those that do not need language and/or country-specific knowledge. Conversely, local platforms will channel those parts of the supply and demand of micro-work which deal with data that cannot be stripped of national or linguistic dimensions. To understand the global micro-working market, we juxtapose them and adopt jointly the national and international lens – thereby complementing and enriching the results of extant literature, which has under-researched national and local platforms so far (Berg et al 2018; Forde et al 2017).

Starting from one country case, our approach enables to explore the linkages between national and international levels in micro-work platform intermediation. It also aims to better define the boundaries of competition by considering not only the services offered but also their geographical reach in terms of the locations of both clients and workers. It is for this reason that we exclude from our selection platforms that do offer AI-for-automobile-services but do not have a presence in France, such as Samasource (USA), Playment (India) and Crowd Guru (Germany). We also exclude professional freelancing platforms like Upwork which focus on offering access to the services of a single provider (not undifferentiated “crowds”) for a specific, usually highly qualified job, such as translation, computer graphics, and software development.

4. How micro-work is crucial to AI production

Let us first go back to the basics, reviewing the essential principles that drive today’s automated systems and connecting them to the need for human inputs. The websites of Mighty AI, Figure Eight, and to a lesser extent IsAHit, address car makers’ perceived willingness to invest strategically into self-driving cars, as if this was the main (or only) application of AI to the automotive industry. It is true that the recent major technological progresses have revamped the long-standing, but previously stagnant, research to develop autonomous vehicles, thereby initiating a fierce race among competitors. But without necessarily replacing human drivers, machines may instead accompany them. Appen (and Gengo, the Lionbridge subsidiary) highlight consumer demand for effective on-board virtual assistants, requiring speech interfaces to access all the functionalities of computers and mobile phones, without ever taking the driver's hands and eyes off the road. Appen also suggests future uses for road safety purposes: if machines learn to recognize drivers' emotions on top of their words, they will predict distracted driving behavior, and forewarn other machines. Lionbridge insists on AI for targeted marketing, using the car infotainment system to suggest products and services, based on data about the driver's preferences and behaviors.

What all these applications have in common, is that at the current level of technological development, they rest on *machine-learning*, a branch of research at the crossroads of informatics and statistics, which develops algorithmic models that access data and use it to figure out a solution to an existing problem (such as sorting information, making a complex decision, interpreting a request). These models are intended to "learn" a solution based on data, without that solution being explicitly programmed. In recent years, machine learning has enabled significant progress in many areas. In autonomous vehicles, for instance, even simple rule-based behaviors such as "slow down if a pedestrian crosses the street" requires recognizing a pedestrian – a trivial task for humans, an extremely complex one for a computer. Machine learning appears as the solution: based on vast amounts of existing images, it can find common patterns (such as shapes and trajectories), and it can use those to learn to distinguish other pedestrians when they appear again on the road. In practice, the algorithm will compare any new passer-by to its stock of images, find similarities, and conclude with some degree of certainty whether it "sees" a pedestrian.

To achieve this, machine learning algorithms need not only large sets of images of pedestrians, but also labels that tell them what these images are – whether they show pedestrians at all, how many of them, where precisely they are positioned in the picture, etc. Therefore, the "raw" images routinely taken by lidars, cameras, and sensor devices mounted on today's robot cars are not enough: to be useful, they first need to be annotated. It is here that micro-workers intervene: platforms send those images to their online micro-task executors, who identify and label everything that can be seen in each of them, from pedestrians and dogs to traffic lights, other cars, bikes etc.

The job is necessarily huge because, as discussed, the machine can only learn from large amounts of data, that would be tedious and lengthy to annotate if just one or few workers were in charge. Rather, platforms fragment these large batches into many short, quick-to-do tasks, and allocate them to many providers, each of whom will do just one or few. In this way, the work can be done much more quickly – and more cheaply in that the platform will just pay for the output received without bearing the full cost of waged labor (which would include benefits, payroll taxes etc.).

What types of tasks do autonomous vehicle producers outsource to micro-workers? The websites of (among others) Figure Eight, Mighty AI and IsAHit offer similar products to potential clients:

- *Image classification*: organizing images by criteria such as quality (detecting blurry pictures, for example), content (what images represent), setting (for example, urban environment vs highway).
- *Object detection or tagging*: identifying traffic objects within images (such as bikes, buses or trees in a street scene) with tools such as bounding boxes, polygons, and outlines.
- *Landmark detection*: pinpointing distinctive features and signs in images.
- *Semantic segmentation*: pixel-level object detection assigning every dot of an image to an object according to a list of relevant types.

These activities all support computer vision and are not specific to the automobile industry (they may be applied, for example, to medical imagery) but vendor platforms often target car producers and provide demos and examples that relate to road traffic. Many of the French workers we surveyed reported having done tasks of these types; in one instance, they annotated images of motocross races where they had to distinguish different parts of the road and describe them in detail.

Schmidt (2019) highlights the increasing complexity of these tasks over time: while even just a few years ago, grossly recognizing a pedestrian or a dog in a street scene was enough, today's clients aim for much higher levels of detail. Full semantic segmentation provides the highest quality, but it is a quite demanding task that requires patience and precision: it may take up to one hour and half to segment just one image accurately. It is for this reason, he says, that clients are reducing their use of generalist platforms such as Mechanical Turk for these tasks, moving toward specialized platforms such as Figure Eight and Mighty AI, which are more expensive but offer better quality control. As mentioned earlier, these specialized services have a structure that emphasizes workforce management and workflow optimization much more than sheer intermediation.

5. Connected work for connected cars

Let us now move to the development of connected cars, where AI assists drivers without providing full automation. To develop effective speech interfaces for human-machine interaction, large audio data sets are required for algorithms to "learn" to recognize sentences and respond adequately. Here, the data need to be not only annotated, but produced in the first place: for each language in which the digital assistant is to be localized, there is a need for examples of different words and sentences to ask about current topics or operational instructions, and these examples must cover a wide range of local accents, vocal range and timbre, and contextual conditions (such as background noise). Micro-working platforms that can mobilize large numbers of workers are much better-suited to provide this diversity, than any single provider.

All the micro-working platforms in our sample propose – regularly or occasionally – tasks that consist in audio-recording one's voice to train a vocal assistant. Several of the Foule Factory workers that we interviewed reported having had to record themselves reading aloud about a dozen short sentences in French; others had to devise and record a few different ways of asking the same question. Most of these workers understood that the goal of these tasks was to train virtual assistants for use in some connected devices such as (though not necessarily) cars.

As with images, machine learning algorithms can rarely be directly applied to the original audio data produced by micro-workers, but require some prior preprocessing, notably annotations allowing the same content to be recognized in different environments and contexts. Further human input is thus necessary to detect pronunciation and accents, assess sentiment, and classify data into categories. The other two areas outlined above – monitoring drivers' behavior and offering targeted advertising – also require production and annotation of data, both visual (for example, face recognition to detect signs of tiredness) and sound (recognition of signals from speech).

According to Schmidt (2019), there is a major difference between data generation tasks such as voice recording, and data annotation tasks (whether for computer vision, speech recognition, or other usages). The former type requires no qualifications except fluency in the demanded language(s), and should ideally be performed by masses of providers, each contributing a small bit, to ensure diversity. Such tasks are therefore more likely to be found in generalist platforms, and to involve individuals who micro-work occasionally or in their spare time. In contrast, data annotation tasks are increasingly complex and demand higher skills, so that they may be executed by smaller numbers of individuals who micro-work regularly or even full time; they are offered more often on specialized platforms (those we defined as "multi-layer" above) and may in future lead to some degree of professionalization of micro-workers.

6. Is micro-work just a passing phase?

One may optimistically believe that growing algorithmic quality will one day dispense with human involvement in providing and preparing "training" datasets. Because algorithms "learn", they will grow better and better even at tasks such as labelling, tagging and segmenting images, or categorizing audio datafiles. Indeed, the challenges of even just five years ago have now become routine tasks, for example distinguishing different types of animals: AI can now easily solve this problem and hardly needs any "cats" or "dogs" labels. But challenges have not disappeared: quite on the contrary, they have moved to the next level, as AI ambitions to address more complex problems. Companies now need customized resources, want to tackle nuances and details, and demand a very high level of precision especially in the automotive sector, where a mistake might be fatal. The open-ended variety of scenes and situations in which vehicles can find themselves is a reason why there is a continuing need for new training data that cover, for example, different lighting and weather conditions at different times of the day and year, as well as road circulation modifications due to construction, special events or accidents. Far from decreasing, the global demand for data generation and especially for data annotation services has thus surged: a recent report by the Cognilytica think-tank (2019) estimates that data preparation tasks represent over 80% of the time consumed in most AI and machine learning projects, and that the market for third-party data labeling solutions is $150M in 2018, growing to over $1B by 2023.

There is another reason why human input will still be needed even with better and better AI solutions: it is the necessity to check the accuracy of those solutions. Among our interviewees, a French micro-worker corrected the automated transcriptions done by a virtual assistant over a period of several months. She had to listen to the original recordings (audio files lasting only few seconds), compare her understanding to the transcript, suggest modifications if needed, and provide feedback as to contextual factors (such as speaker's accent and fluency, or background noise). The virtual assistant she was working for was far from a prototype still in the "training" phase: it was a widely-sold product in commerce, though currently not for use in cars. The important lesson learned is that even when AI solutions are mature (as will happen at some point in the automotive industry), they still require human help to fine-tune their responses and improve their performance.

The concept of "human-in-the-loop" often employed in computer science, refers to production processes in which both machines and people play a part, at different stages – not just at input but also at output levels. However, except for Figure Eight (whose "8" symbolizes the loop), the platforms considered in this study tend to communicate more on preparation of training data, while busily playing down the use of micro-work for algorithmic checking and for verification of AI solutions. Yet, they all offer this type of tasks, which according to Schmidt (2019) are going to be in growing demand in the years to come. Cognilytica (2019) suggests the same conclusion by considering that corporate demand is going up at both input and output levels: "The human in the loop is not going away any time soon *for data labeling and AI quality control*" (emphasis added).

To sum up, micro-work is not bound to disappear with the development of AI solutions. Instead, its role and relevance will likely increase. What is changing is the type of tasks offered, relatively more diverse and more complex than those classically posted on Amazon Mechanical Turk. As mentioned above, Schmidt (2019) also considers that there is a shift from generalist to more specialized platforms. In addition to confirming these findings, our work stresses the importance of platform diversification, with both global and local channels for AI services based on micro-work – as we are going to discuss in what follows.

7. Diversification within the crowd

It is well-known that large portions of online work, demanded by companies in North America and in Europe, are outsourced to emerging and low-income economies especially in South Asia (Graham et al. 2017; Lehdonvirta et al. 2019). While extant literature has focused on English-speaking countries, the same is likely to be happening in other linguistic regions (Casilli 2017). While we do not yet have enough data to accurately map this phenomenon, we have gathered circumstantial evidence by meeting French entrepreneurs who talked about outsourcing to pools of micro-workers in French-speaking African countries. Moreover, we have interviewed a selected set of micro-workers who execute tasks for French clients from Cameroon, the Ivory Coast and Madagascar. Clearly, outsourcing to Africa is an economically viable choice insofar as most data-production and data-annotation tasks can be executed anywhere, requiring only a computer and broadband connection.

And yet, not all micro-work is done in countries where labor costs are low. In our fieldwork, we have interviewed and observed several micro-workers based in France, some of whom engage in online labor regularly, including a few for whom it is the sole source of earnings. Our own estimate of the French country-level micro-workforce accounts for over 50,000 routine workers and up to 250,000 occasional ones (Le Ludec et al. 2019). Significantly, the Foule Factory platform bases its business model on recruiting French providers, and it recommends paying the French minimum wage to them. So, is global competition limited after all, and if so, what are the barriers to it?

The reason is that there is a need for niche competencies, resulting in lower substitutability among micro-workers. Reading aloud sentences in French is a task that only native speakers can properly do; so is ability to detect accents, fluency or pronunciation errors. Similarly, recognition of objects such as French license plates or traffic signs is a task that requires some familiarity with the local context, local experience or at least some *ad hoc* training. In short, it is because language and culture matter (Irani 2015), that local platforms are relevant, and the study of global patterns in micro-work cannot be based only on international, English-speaking platforms. It is also for this reason that, even within the same linguistic area, not all tasks can be outsourced to workers in the low-income world, and residents of industrial countries like France also micro-work.

8. Conclusions: BPA as BPO, and plans for ethical AI

The above considerations suggest that what goes under the name of Business Process Automation (BPA) is in fact often Business Process Outsourcing (BPO). Humans are involved at different stages of the automation process, notably in the car industry, and with current trends in technological development, there are no signs that their role is going to diminish. Popular fears that "robots will steal our jobs" forget the crowds who silently toil to allow robots to function in the first place. The issue is not the technical allocation of tasks to humans and machines *per se*, but rather the underlying distribution of power and the design of the organizational set-ups that support the production and deployment of technologies.

The micro-workers that make AI possible are usually external to the AI-developing technology firm, formally recruited as independent contractors through a dedicated online platform. Anonymous, largely interchangeable, and removed from the core business, they remain out of sight and have little opportunity to connect with (and sometimes, even be aware of) fellow workers. Even when platforms or clients acknowledge the collective effort to train and calibrate artificial intelligence technologies, they frame it in a rhetoric of ephemeral and devalued work that does not need to be advertised as such – what matters is the final product, without much emphasis on its origins. This lack of both peer-based and exogenous recognition is, to date, the main obstacle to the emergence of a new subjectivity predicated on platform labor (Casilli 2019).

This is one effect that the surge in AI production is already exerting on labor markets, even though it is not yet fully deployed in many markets such as the automobile. Questions about the effects of AI should therefore take notice of these precociously-affected workers, the conditions under which they operate, the remunerations they receive, and the future perspectives open to them. Proper governance of AI and the development of ethical rules for it, widely discussed today, require full and explicit consideration of what happens in its back-office.

9. Acknowledgements

Data used in this study are from the DiPLab ("Digital Platform Labor") research project, co-funded by Maison des Sciences de l'Homme Paris-Saclay; Force Ouvrière, a workers' union, through IRES (Social and Economic Research Institute); and France Stratégie, a service of the French Prime Minister. We also thank Foule Factory and IsAHit for logistical support, and Inria for complementary funding.

10. Conflicts of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

11. References

Autor, D.H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, 29(3): 3–30.

Bastin, G. and Tubaro, P. (2018). Le moment big data des sciences sociales. *Revue française de sociologie*, 59(3): 375-394.

Braverman, H. (1974). *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century*. New York, Monthly Review Press.

Brynjolfsson, E. and McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York, Norton.

Casilli, A.A. (2017). Digital labor studies go global: Toward a digital decolonial turn. *International Journal of Communication*, 11: 3934–3954.

Casilli, A.A. (2019). *En attendant les robots. Enquête sur le travail du clic*. Paris, Editions du Seuil.

Cognilytica Research (2019). *Data Engineering, Preparation, and Labeling for AI*. Report.

Difallah, D., Filatova, E. and Ipeirotis, P.G. (2018). Demographics and dynamics of Mechanical Turk workers. In *Proceedings of WSDM 2018: the Eleventh ACM International Conference on Web Search and Data Mining*. ACM, 135–143.

Ekbia, H. and Nardi, B. (2017). *Heteromation, and Other Stories of Computing and Capitalism*. Cambridge (MA), MIT Press.

Forde, C., Stuart, M. Joyce, S., Oliver, L., Valizade, D., Alberti, G., Hardy, K., Trappmann, V., Umney, C. and Carson, C. (2017). *The Social Protection of Workers in the Collaborative Economy*. Report for European Parliament Employment and Social Affairs Committee.

Frey, C. and Osborne, M. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114(C): 254–280.

Fuchs, C. (2019). Karl Marx in the age of big data capitalism. In *Digital Objects, Digital Subjects: Interdisciplinary Perspectives on Capitalism, Labour and Politics in the Age of Big Data*, D. Chandler and C. Fuchs (eds.). London, University of Westminster Press: 53-71.

Graham, M., Hjorth, I., and Lehdonvirta, V. (2017). Digital labour and development: impacts of global digital labour platforms and the gig economy on worker livelihoods. *Transfer: European Review of Labour and Research*, 23(2): 135–162.

Gray, M. and Suri, S. (2017). The humans working behind the AI curtain. *Harvard Business Review*, (9): 2–5.

Irani, L. (2015). Difference and dependence among digital workers: The case of Amazon Mechanical Turk. *South Atlantic Quarterly*, 114.

Irani, L. (2016). The labor that makes AI "magic". *AINow Summit*, White House and NYU.

Kuek, S., Paradi-Guilford, C., Fayomi, T., Imaizumi, S., and Ipeirotis, P. (2015). *The global opportunity in online outsourcing*. Report, World Bank.

Lehdonvirta, V., Kässi, O., Hjorth, I., Barnard, H. and Graham, M. (2019). The global platform economy: A new offshoring institution enabling emerging-economy microproviders. *Journal of Management*, 45(2): 567-599.

Le Ludec, C., Tubaro, P. and Casilli, A. (2019). *How many people microwork in France? Estimating the size of a new labor force*. Working paper, arXiv:1901.03889 [econ.GN].

Ohnemus, J., Erdsiek, D. and Viete, S. (2016). *Nutzung von Crowdworking durch Unternehmen: Ergebnisse einer ZEW-Unternehmensbefragung*. Report, BMAS Forschungsbericht 473. Bundesministerium für Arbeit und Soziales.

Ricardo, D. (1951 [1821]), *On the Principles of Political Economy and Taxation*, vol. 1 of *Works of David Ricardo,* P. Sraffa and M.H. Dobb (eds.). Cambridge, Cambridge University Press.

Rochet, J. and Tirole, J. (2003). Platform competition in two-sided markets*. Journal of the European Economic Association*, 1: 990-1029.

Schmidt, F.A. (2019). *Crowdproduktion von Trainingsdaten: Zur Rolle von Online-Arbeit beim Trainieren autonomer Fahrzeuge.* Report, Hans-Böckler-Stiftung.

Star, S.L. and Strauss, A. (1999). Layers of silence, arenas of voice: The ecology of visible and invisible work. *Computer Supported Cooperative Work*, 8(1-2): 9-30.